Healthcare Twitter Analytics

Do any of the numerics predict score?

```
file = 'Tweets_Celiac_sent.csv'
data = read.csv(file)
data=data[c("listed_count","favourites_count","sentiment","retweet_count","score")]
model1 = lm(score~.,data=data)
summary(model1)
```

```
##
## Call:
## lm(formula = score ~ ., data = data)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -9.392 -1.029 -0.656  0.503  5.146
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)      7.92e+00   2.47e-02  320.14  < 2e-16 ***
## listed_count     4.91e-04   3.25e-05   15.13  < 2e-16 ***
## favourites_count 8.38e-06   3.58e-06    2.34  0.01937 *
## sentiment        3.26e-02   8.62e-03    3.79  0.00015 ***
## retweet_count    7.15e-03   5.00e-03    1.43  0.15254
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.51 on 5670 degrees of freedom
## Multiple R-squared:  0.0448, Adjusted R-squared:  0.0441
## F-statistic: 66.4 on 4 and 5670 DF,  p-value: <2e-16
```

```
model2 = lm(score~.-retweet_count,data=data)
summary(model2)
```

```
##
## Call:
## lm(formula = score ~ . - retweet_count, data = data)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -9.436 -1.022 -0.658  0.502  5.291
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)      7.93e+00   2.32e-02  341.31  < 2e-16 ***
## listed_count     4.99e-04   3.21e-05   15.55  < 2e-16 ***
## favourites_count 8.78e-06   3.57e-06    2.46  0.01410 *
## sentiment        3.22e-02   8.61e-03    3.74  0.00019 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.51 on 5671 degrees of freedom
## Multiple R-squared:  0.0444, Adjusted R-squared:  0.0439
## F-statistic: 87.9 on 3 and 5671 DF,  p-value: <2e-16
```

```
lm_compare(model1, model2)
```

```
## Residual Standard Error
##       model1          1.514
##       model2          1.514
##       Increased:      0.0001398
##               model1 preferred
## Adjusted R^2
##       model1          0.04409
##       model2          0.04391
##       Decreased:      -0.0001765
##               model1 preferred
## F Statistic
##       model1          66.42
##       model2          87.86
##       Increased:      21.44
##               model2 preferred
## F Statistic p-value
##       model1          0
##       model2          0
##       Unchanged:      0
##
## Coeffcient Statistics
##   listed_count  abs(t stat)
##       model1          15.13
##       model2          15.55
##       Increased:      0.4163
##               model2 preferred
##   listed_count  t stat p-value
##       model1          1.01e-50
##       model2          2.14e-53
##       Decreased:      -1.008e-50
```